OLASILIK TEORİSİ VE İSTATİSTİK

ÖRNEKLEME VE ÖRNEKLEM DAĞILIMLARI

- İstatistikte temel amaç, kitlenin tümü hakkında bilgi edinmektir. Önceki konularda da gördüğümüz gibi incelenilen toplumun tümü **kitle (ana kitle)** olarak isimlendirilmişti.
- Ancak, kitleye ulaşmanın çoğu zaman olanaksız olması nedeniyle bu kitleyi en iyi şekilde temsil edecek bir alt gruptan yararlanmak, istatistiğin başlıca konusudur.
- Kullanılan bu alt gruba da **örneklem** denilmişti. Bazı örnekler vererek kitleyi hatırlayalım:

- Bir sigorta şirketinde sigortalı araçlardan kaza yapıp sigortadan masrafları alanların tümü
- İhracat yapan bir firmanın bir yıllık tüm ihraç kalemleri
- Ankara'da yaşayan ailelerde yapılacak bir gelir araştırmasında Ankara'da yaşayan ailelerin tümü
- Bir banka hesaplarının kontrolünde tüm hesap kalemleri

gibi örnekler verilebilir.

Yukarıda verilen örneklerden kitleye ait bilgi almanın güçlüğü açıktır.

- Gerek zaman gerek maliyet kısıtı birer engel olarak karşımıza çıkar.

- Bu nedenle örneklemden alınan bilgiden yararlanılır.

- Ancak seçimi iyi yapılmamış bir örneklem de bizi yanıltır.
- Örneğin, yeni bir gıda maddesi üretmiş olan bir firma bunun tutunup tutunamayacağını çevresindeki kişilerden aldığı bilgilerle kararlaştırırsa, elbette hatalı karar vermiş olacaktır.
- Kitleden örneklem çekerek bilgile elde etmekteki asıl amaç, kitleye ait doğru bilgiye ulaşmaktır.

- Doğru bilgi elde etmenin yolu doğru örneklemeden geçer. Örnekleme süreci rasgelelik ilkesini uygulamakla sağlanır.
- Rasgelelik uygulamanın amacı kitledeki her bireye eşit seçilme hakkı tanımaktır.

- Rasgelelik   ilkeleri kitlenin yapısına göre değişir.
- En basit örnekleme şekli **basit rasgele örneklemedir.**
- Basit rasgele örneklemede kitlenin her noktası aynı özelliği taşır.

- Yukarıda verilen banka hesaplarının kontrolünde tüm hesap kalemlerinin oluşturduğu kitlede dosyalar kitle birimidirler ve aynı özelliği taşırlar.
- Bu kitle için basit rasgele örnekleme en uygun örnekleme şeklidir.
- Basit rasgele örnekleme, en yaygın kullanılan örnekleme şeklidir.
- Ancak her kitle için uygun olduğu söylenemez.

- Rasgele seçim şeklinin çeşitli yolları vardır.
- Basit rasgele örnekleme ile gerçekleştirilen örnekleme yöntemi ile elde ettiğimiz örneklemlerden yararlanıp, kitle hakkında bilgi elde edilmeye çalışılacaktır.

- Kitleye özgü niteliklere **parametre,** örnekleme özgü niteliklere **istatistik** demiştik. Bu iki kavram birbirlerinden farklıdır**.**

- Yukarıda verilen dördüncü örnekte tüm dosyalardaki hata oranı parametredir.
- Bu dosyalar içinden rasgelelik kurallarına uyularak çekilen elli birimlik örneklemden hata oranları elde edilir ise istatistik elde edilmiş olur.
- Kitleye ait hata oranı sabit bir değerdir.
- Ancak, genellikle bilinemez.

- Örnekleme ait hata oranı seçilen örneklemden örnekleme değişir.
- Bu bilgilerin bir dağılımı olacaktır bu dağılıma **örneklem dağılımı denir**.
- Bu bölümün amacı **örneklem dağılımlarının** özelliklerini incelemektir.

- Yukarıda verilen örnekte örneklem dağılımı, oranlara ait idi.
- Örneklem dağılımları ortalama, varyans ya da örneklemin diğer özellikleri için düşünülebilir.

- Örnek 1 örneklem ortalamasının örneklem dağılımını incelemektedir.

- **Örnek 1.** Bir iş yerinde altı işçi çalışmaktadır. Bu işçilerden dört kişilik gruplar oluşturulup vardiyalı çalıştırılması istenmektedir. İşçilerin iş deneyimleri yıl olarak aşağıda verilmiştir:
- 2 4 6 6 7 8
- Bu altı işçiden dörtlü gruplar 15 farklı şekilde seçilebilir. Çizelge1'de bu seçimler verilmiştir.
- Bu örnekte kitle çok küçüktür ve oluşturulabilecek örneklem sayısı da 15 gibi az bir sayıdır.
- Uygulamada bu kadar az sayıda örneklem oluşturulması mümkün olmaz.
- Bu küçük örnek, konuyu örnek ile açıklayabilmek için verilmiştir.
- Elimizdeki kitlenin ortalaması

  $(2+4+6+6+7+8)/6 = 5.5 = \mu x$ (parametre)

- **Çizelge 1.** 2,4,6,6,7,8 Kitlesinden Çekilen Dört Birimlik Örneklemler ve Ortalamaları

ÖRNEKLEM	ORTALAMASI	ÖRNEKLEM	ORTALAMASI
2,4,6,6	4.50	2,6,7,8	5.75
2,4,6,7	4.75	2,6,7,8	5.75
2,4,6,8	5.00	4,6,6,7	5.75
2,4,6,7	4.75	4,6,6,8	6.00
2,4,6,8	5.00	4,6,7,8	6.25
2,4,7,8	5.25	4,6,7,8	6.25
2,6,6,7	5.25	6,6,7,8	6.75
2,6,6,8	5.50		

- Not: Bu örnek P.Newbold İşletme ve İktisat İçin İstatistik ss.246'dan alınmıştır.

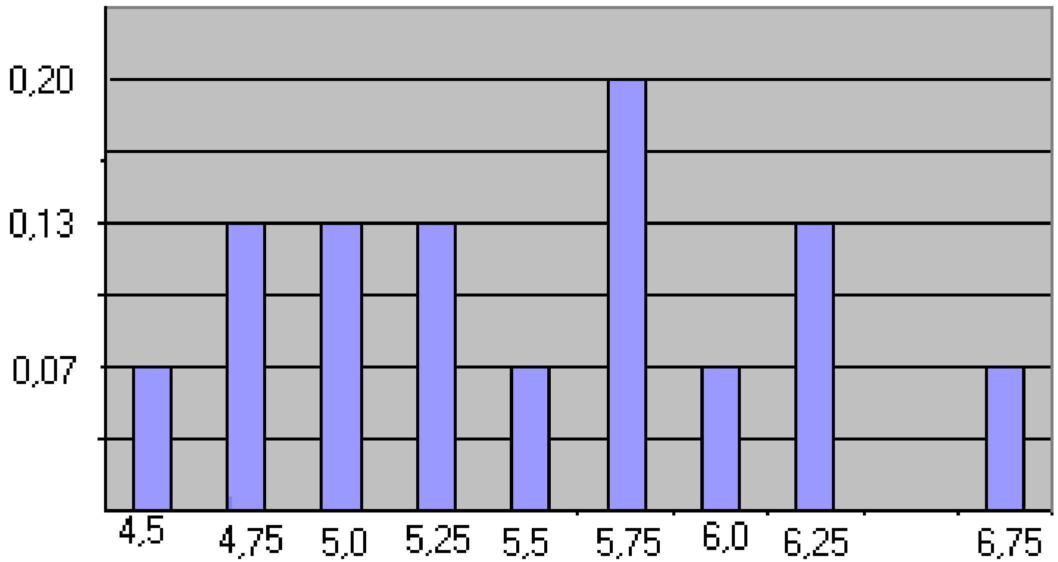
- Çizelge 1 incelendiğin de, kitleden çekilen örneklemlerin ortalamalarının farklı olduğu ve bir dağılımının olacağı görülür.

- Ortalaması 5.75 değerini alan 3 örneklem,
- ortalaması 4.75 değerini alan 2 örneklem,
- ortalaması 6.00 değerini alan 1 örneklem,
- ortalaması 5.00 değerini alan 2 örneklem,
- ortalaması 6.25 değerini alan 2 örneklem,
- ortalaması 5.55 değerini alan 2 örneklem,
- ortalaması 6.75 değerini alan 1 örneklem,
- ortalaması 5.50 değerini alan 1 örneklem bulunmaktadır.

- Ortalamaların almış olduğu bu değerler olasılık foksiyonu olarak ifade edilirse:

- P($\overline{X}$ =5.75)=3/15 ; P($\overline{X}$ =4.75)=2/15 ; P($\overline{X}$= 6.00)=1/15 ;
- P($\overline{X}$=5.00)=2/15 ; P($\overline{X}$ =6.25)=2/15 ; P($\overline{X}$ =5.55)=2/15 ;
- P($\overline{X}$ =6.75)=1/15 ; P($\overline{X}$ =5.50)=1/15 olur.

• Ortalamaların olasılık dağılımına ilişkin dağılım grafiği aşağıda verilmiştir.



Örneklem Ortalamasının Örneklem Dağılımı

- Birinci bölümde örneklem dağılımlarından bahsedildi ve ortalamalara ait bir örneklem dağılımı örneği verildi.
- Bu bölümde örneklem ortalamasının dağılımından daha ayrıntılı olarak bahsedeceğiz.
- Ortalaması μ_x, varyansı σ^2_x olan bir kitleden gözlemleri $x_1, x_2, \ldots, x_n$ olan n büyüklüğünde bir örneklem çekildiği düşünülsün.
- Bu örneklemin ortalamasının

$$\overline{X} = \sum_{i=1}^{n} x_i / n$$

- Burada $\overline{X}$ nın örneklem dağılımını inceleyeceğiz. Bir dağılımın öncelikli bilgisinin ortalama sonra varyans olduğunu hatırlayalım. Bir dağılımın ortalamasının raslantı değişkenlerinin beklenen değerleri alınarak bulunduğu da bilinmektedir. Bu durumda E($\overline{X}$) bulunması gerekir. Önceki bilgilerimizden aşağıda verilen eşitlik yazılabilir:

$$E(\overline{X}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)$$

- Yukarıda verilen eşitlikte beklenen değer kurallarını hatırlayalım.
- 1/n sabit bir değerdir ve beklenen değeri kendisidir. Toplamların beklenen değeri de beklenen değerler toplamına eşittir.
- Bu kurallar uygulanarak yukarıda verilen eşitliğin sağ tarafı aşağıda verildiği şekilde

$$E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}[E(x_1) + E(x_2) + \cdots + E(x_n)]$$

- Burada her x_i ortalaması μ_x, varyansı $\sigma^2{}_x$ olan bir kitleden çekilmiş bir raslantı değişkeni olduğu için beklenen değerleri μ_x olacaktır. Yukarıda verilen eşitliğin sağ tarafı $\frac{1}{n}$ n μ_x olur. Sonuç olarak;

- $E(\overline{X})=\mu_x$

- elde edilir.

- Bu sonuçtan çıkacak anlam; Ortalaması μ_x olan bir kitleden ard arda örneklemler çekilse bu örneklemlerin ortalamalarının ortalaması, kitle ortalamasına yakınsar.
- **Örnek 2.** Örnek1 de verilen problemde elde edilen sonucun doğruluğunu gösterelim.
- Problemimiz; bir iş yerinde altı işçi çalışmaktadır. Bu işçilerden dört kişilik gruplar oluşturulup vardiyalı çalıştırılması istenmektedir.
- İşçilerin iş deneyimleri yıl olarak
- 2 4 6 6 7 8
- şeklinde idi.

- Bu altı işçiden dörtlü gruplar 15 farklı şekilde seçilebileceği de gösterilmişti ve bunlara ait ortalamalar
- $\bar{x}_1$=4.50; $\bar{x}_2$ =4.75 $\bar{x}_3$=5.00; $\bar{x}_4$ = 4.75; $\bar{x}_5$ =5.00 ;

$\bar{x}_6$ =5.25; $\bar{x}_7$ =5.25; $\bar{x}_8$ =5.50 ; $\bar{x}_9$ =5.75; $\bar{x}_{10}$ =5.75;

- $\bar{x}_{11}$ =5.75; $\bar{x}_{12}$ =6.00; $\bar{x}_{13}$ =6.25; $\bar{x}_{14}$ =6.25;
- $\bar{x}_{15}$ =6.75 olarak bulunmuştu.
- Bu ortalamalar kitlemizde oluşturulabilecek tüm örneklemlere ait olup ortalamaları;
- (4.50+4.75+5.00+4.75+5.00+5.25+5.25+5.50+5.75+5.75 +5.75+6.00+6.25+6.25+6.75)/15=5.5

- Aynı sonuç kesikli raslantı değişkenlerinde ortalama için verilen
- $E(\overline{X}) = \sum \overline{x} P_{\overline{x}}(x)$ eşitlikten de elde edilir;

- P($\overline{X}$ =5.75)=3/15; P($\overline{X}$ =4.75)= 2/15 ; P($\overline{X}$ = 6.00)=1/15 ; P($\overline{X}$ =5.00)=2/15 P($\overline{X}$ =6.25)=2/15; P($\overline{X}$ =5.55)=2/15 ; P($\overline{X}$ =6.75)=1/15 ; P($\overline{X}$ =5.50)=1/15; P($\overline{X}$ =4.50)=1/15
- kullanılarak,

$E(\overline{X}) = \sum \overline{x} P_{\overline{x}}(x)$ =4.50(1/15)+4.75(2/15)+5.00(2/15)+…+6.75(1/15)=5.5

- kitle ortalamasını elde edebiliriz. Her zaman bu örnekte olduğu gibi oluşabilecek tüm örneklemler oluşturulamayacağı için n büyüdükçe örneklem ortalamasının kitle ortalamasına yakınsadığını bu örneğe göre söyleyebiliriz..

- # Referanslar

- Z. Muluk, S.Cula, «Olasılık Teorisi ve İstatistik Ders notları», Başkent Üniversitesi.
- Sheldon Ross, "Olasılık ve İstatistiğe Giriş (Mühendisler ve Fenciler için), Nobel Yayın Dağıtım, 2010.
- Sheldon Ross, "A First Course in Probability", 7,8 veya 9. baskı, Pearson
- D.C. Montgomery, G.C. Runger, "Applied Statistics and Probability for Engineers", 5th Edition, Wiley, 2011, ISBN: 9780470505786.